Supplemental Materials

Revisiting DNA Sequence-dependent Deformability in High-resolution Structures: Effects of Flanking Base Pairs on Dinucleotide Morphology and Global Chain Configuration

R.T. Young[1], L. Czapla[1,2], Z.O. Wefers[1], B.M. Cohen[1] and W.K. Olson[1*]

1 Department of Chemistry & Chemical Biology, Center for Quantitative Biology, Rutgers, the State University of New Jersey, Piscataway, New Jersey 08854, USA

2 Memorial Sloan Kettering Cancer Center, New York City, New York 10065, USA

* Correspondence to Wilma K. Olson (wilma.olson@rutgers.edu**).**

Overview

The contents of the Supplemental Materials include a figure illustrating the organization of DNA tetramer sequences used in this work (Figure S1); a table listing all protein-DNA structures collected for this study (Table S1); tables with the counts, twist, and volume features of the 10 unique DNA with respect to year (Table S2) and resolution (Table S3); tables with the counts, twist, and volume features of all 136 unique tetramers and their associated dimers for the entire dataset (Table S4) and for all structures at 3.0-Å or better resolution (Table S5); a table detailing sequence acronyms used in the text and the previously-reported ratios of apparent to real chain length (Table S6). All tables can be found in the accompanying .xls spreadsheet document with each table in a separate worksheet.

	AA	AG	AC	AT	GA	GG	GC	GT	CA	CG	CC	CT	TA	TG	TC	TT
AA	AAAA	AAAG	AAAC	AAAT	AAGA	AAGG	AAGC	AAGT	AACA	AACG	AACC	AACT	AATA	AATG	AATC	AATT
GA	GAAA	GAAG	GAAC	GAAT	GAGA	GAGG	GAGC	GAGT	GACA	GACG	GACC	GACT	GATA	GATG	GATC	GATT
CA	CAAA	CAAG	CAAC	CAAT	CAGA	CAGG	CAGC	CAGT	CACA	CACG	CACC	CACT	CATA	CATG	CATC	CATT
TA	TAAA	TAAG	TAAC	TAAT	TAGA	TAGG	TAGC	TAGT	TACA	TACG	TACC	TACT	TATA	TATG	TATC	TATT
AG	AGAA	AGAG	AGAC	AGAT	AGGA	AGGG	AGGC	AGGT	AGCA	AGCG	AGCC	AGCT	AGTA	AGTG	AGTC	AGTT
GG	GGAA	GGAG	GGAC	GGAT	GGGA	GGGG	GGGC	GGGT	GGCA	GGCG	GGCC	GGCT	GGTA	GGTG	GGTC	GGTT
CG	CGAA	CGAG	CGAC	CGAT	CGGA	CGGG	CGGC	CGGT	CGCA	CGCG	CGCC	CGCT	CGTA	CGTG	CGTC	CGTT
TG	TGAA	TGAG	TGAC	TGAT	TGGA	TGGG	TGGC	TGGT	TGCA	TGCG	TGCC	TGCT	TGTA	TGTG	TGTC	TGTT
AC	ACAA	ACAG	ACAC	ACAT	ACGA	ACGG	ACGC	ACGT	ACCA	ACCG	ACCC	ACCT	ACTA	ACTG	ACTC	ACTT
GC	GCAA	GCAG	GCAC	GCAT	GCGA	GCGG	GCGC	GCGT	GCCA	GCCG	GCCC	GCCT	GCTA	GCTG	GCTC	GCTT
CC	CCAA	CCAG	CCAC	CCAT	CCGA	CCGG	CCGC	CCGT	CCCA	CCCG	CCCC	CCCT	CCTA	CCTG	CCTC	CCTT
TC	TCAA	TCAG	TCAC	TCAT	TCGA	TCGG	TCGC	TCGT	TCCA	TCCG	TCCC	TCCT	TCTA	TCTG	TCTC	TCTT
AT	ATAA	ATAG	ATAC	ATAT	ATGA	ATGG	ATGC	ATGT	ATCA	ATCG	ATCC	ATCT	ATTA	ATTG	ATTC	ATTT
GT	GTAA	GTAG	GTAC	GTAT	GTGA	GTGG	GTGC	GTGT	GTCA	GTCG	GTCC	GTCT	GTTA	GTTG	GTTC	GTTT
CT	CTAA	CTAG	CTAC	CTAT	CTGA	CTGG	CTGC	CTGT	CTCA	CTCG	CTCC	CTCT	CTTA	CTTG	CTTC	CTTT
TT	TTAA	TTAG	TTAC	TTAT	TTGA	TTGG	TTGC	TTGT	TTCA	TTCG	TTCC	TTCT	TTTA	TTTG	TTTC	TTTT

Figure S1. Organization of DNA tetramer sequences used in the current work. Entries above the diagonal, depicted in black font, correspond to the 136 unique tetramers, and those below the diagonal, shown in white font, to the complementary sequences. Data are organized such that purine-purine steps lie at the top left of the grid (blue), purine-pyrimidine steps at the lower left (green), pyrimidine-purine steps at the top right (gold), and pyrimidine-pyrimidine steps at the lower right (blue). Labels along the vertical axis, corresponding to the first two bases, are combined with those along the top horizontal axis, corresponding to the last two base pairs, to generate the listed sequences.